What are the Chances? A Look at Bayesian Network Modeling as a Predictive Decision Support Tool

PRESENTED AT THE 11TH NATIONAL MONITORING CONFERENCE

MARCH 25-29, 2019

DENVER, CO

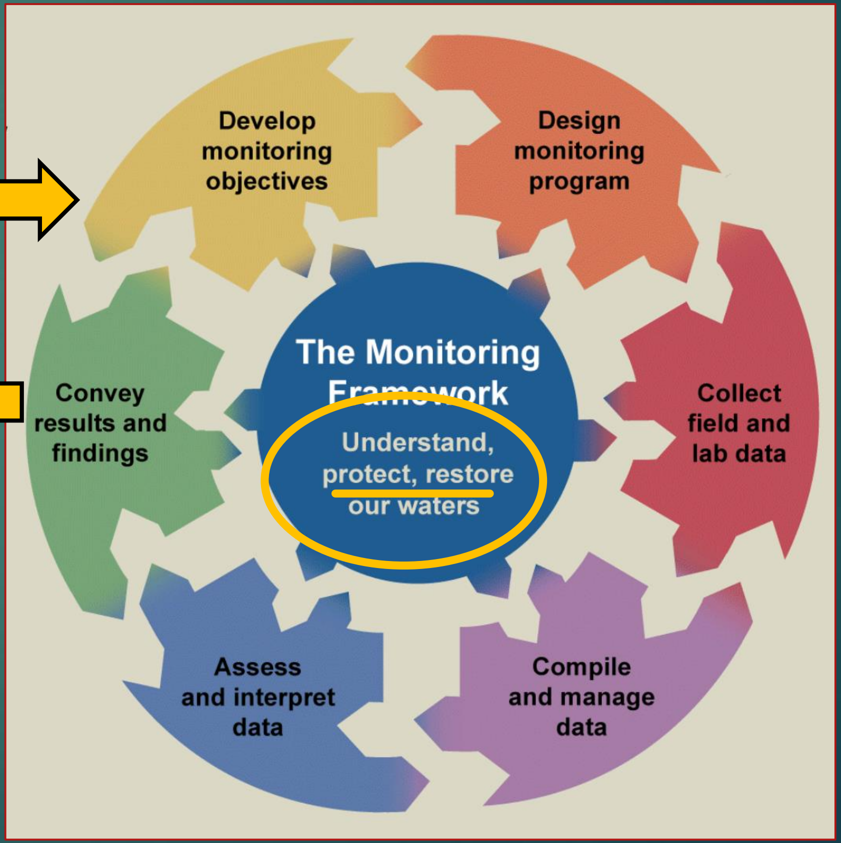
DOUGLAS MCLAUGHLIN, KIESER & ASSOCIATES, LLC

KIESER & ASSOCIATES
ENVIRONMENTAL SCIENCE & ENGINEERING

Decision-making & the NWQMC Monitoring Framework

It's About the Decision

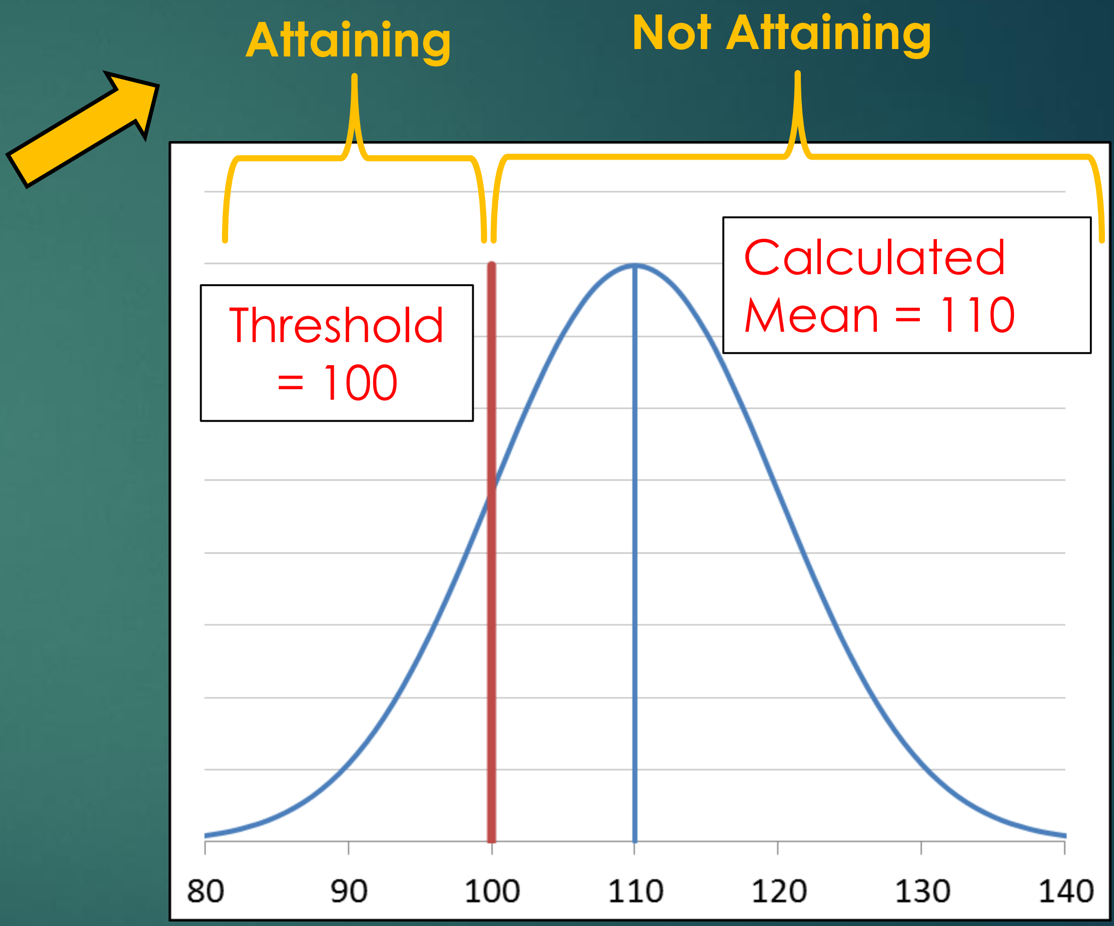
Maximizing the chances of taking cost-effective **<u>action</u>**



How to best assess, interpret, and "convey results and findings" to decision-makers?

Water Resources Decisions Often Involve…

- Numeric Thresholds
  - Often create two or three **categories** or "conditions"
    - acceptable/not acceptable
    - good/fair/poor

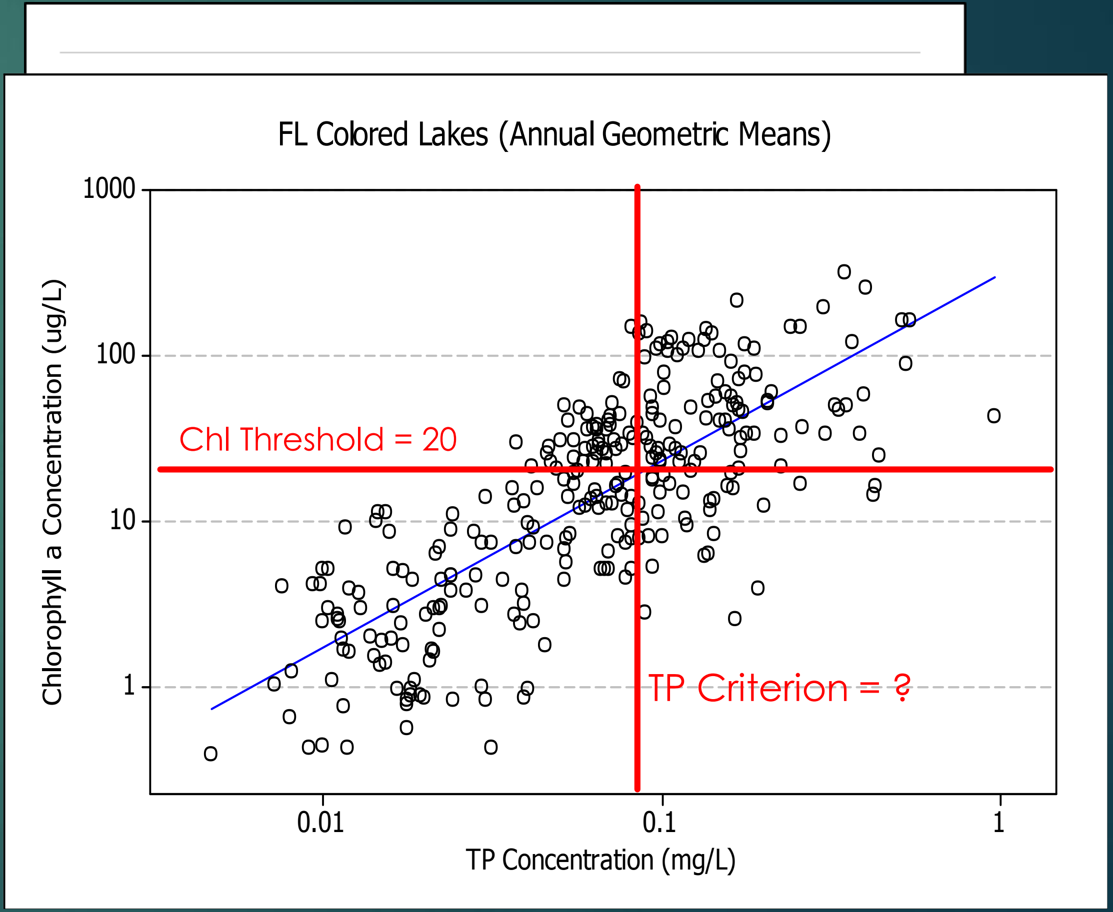


Water Resources Decisions Often Involve…

- Numeric Thresholds
  - Often create two or three categories or "conditions"
    - acceptable/not acceptable
    - good/fair/poor

And

- Causal Relationships
  - How to achieve desired EPT status?
    - e.g., goal is EPT > 14
  - How to support conditions for coldwater species?
    - e.g.,goal is coldwater macroinvertebrate taxa >= 4
  - How to minimize excessive algal growth?
    - e.g., goal is chl *a* < 20 ug/L



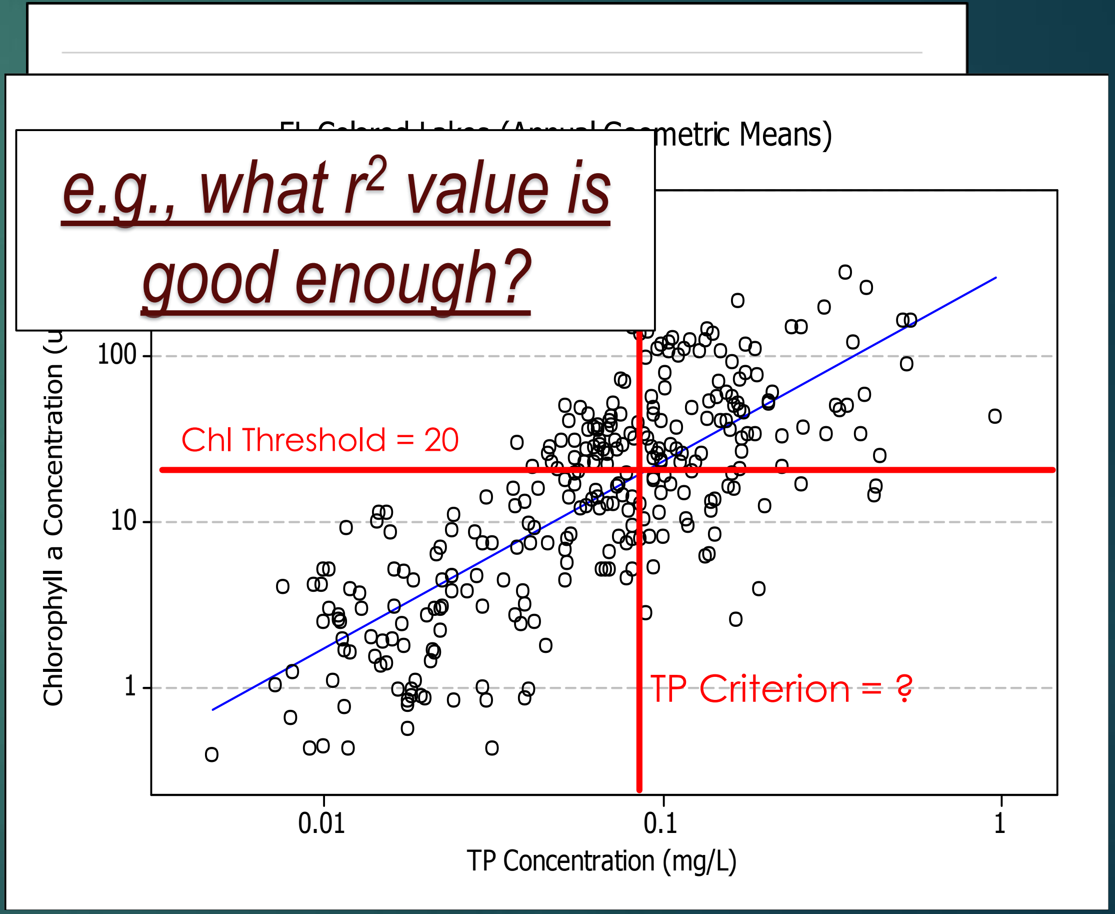
Wrestling With Uncertainty

How to measure it. Its Impact on decision-making. How to communicate it.

Confidence In

It's About the Decision

e.g., what r^2 value is good enough?

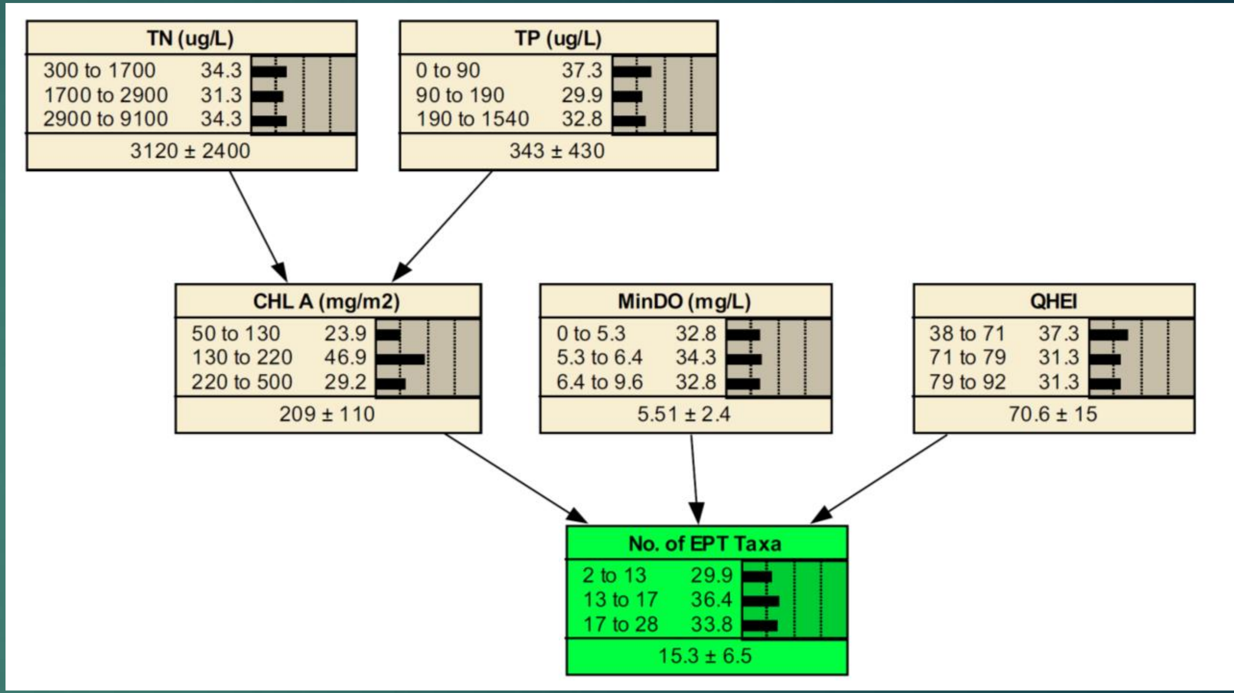


Common Questions for Decision Makers

Common Management Questions
What's our goal, and are we meeting it?
If not, what action should we take?
How confident are we that the system will respond as we think?
Can stakeholders clearly see what actions are being recommended and why?

What Is A Bayesian Network Model?

- Bayes Theorem
- Conceptual models/influence diagrams
- Nodes (parent, child) and links showing relationships
- Bins/states, bin boundaries can reflect useful categories, thresholds
- Conditional probability tables (CPT) describing likelihoods of an outcome state given combinations of parent nodes
- Output in terms of probabilities (i.e., "what are the chances" of observing an output state/bin?)



From McLaughlin & Reckhow (2017)

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

Example: A = EPT taxa; B = QHEI

A really helpful paper about BN models: Marcot et al. (2006)

“Like A Drunk Uses A Lightpost…”

- “Statistics should be used like a drunk uses a lightpost…more for support than illumination.” *(or something like that. Dominique DiToro, at a conference sometime in the late ‘80s)*
- The application here: BNs, like other modeling procedures, should be supported by a range of other analyses, including constructing a solid conceptual model and simply “poring over the data.”

Table 2
Spearman rank correlation coefficient for BN model endpoints (bold values $p < 0.05$).

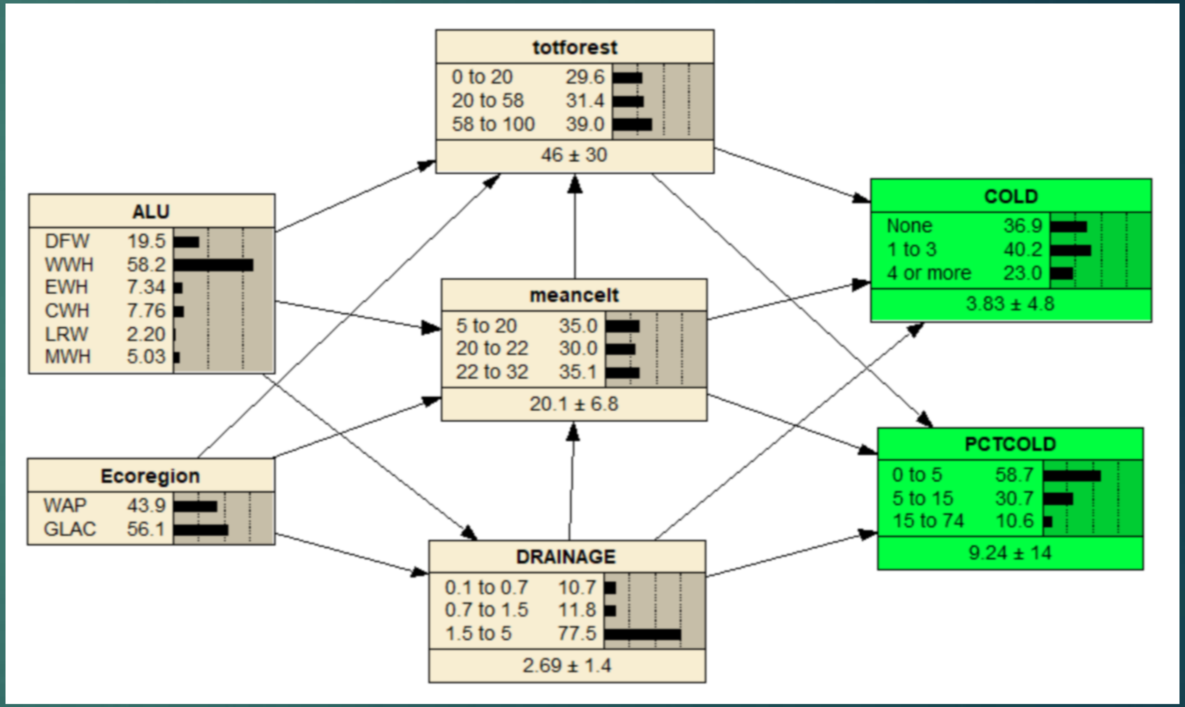
	EPT	CHL	QHEI	MinDO	TP	TN
CHL	**0.34**					
QHEI	**0.48**	0.09				
MinDO	**0.48**	−0.02	0.16			
TP	**−0.39**	**−0.25**	−0.10	−0.19		
TN	0.16	0.00	0.11	**0.31**	**0.56**	
TKN	**−0.62**	**−0.36**	**−0.25**	**−0.29**	**0.81**	**0.36**

From McLaughlin & Reckhow (2017)

BN Coldwater Macroinvertebrate Model (Ohio)

- Percent forest cover, mean water temperature, drainage area as predictors of coldwater macroinvertebrate taxa richness ≥4 in two ecoregions in Ohio
- CPTs generated using field data (954 records, expectation-maximization learning algorithm)
- Selection of explanatory variables and bin boundaries supported by other information and analyses
- One of the objectives: Provide a screening tool for identifying coldwater aquatic life use defined as coldwater taxa richness ≥4.

Calibrated Model

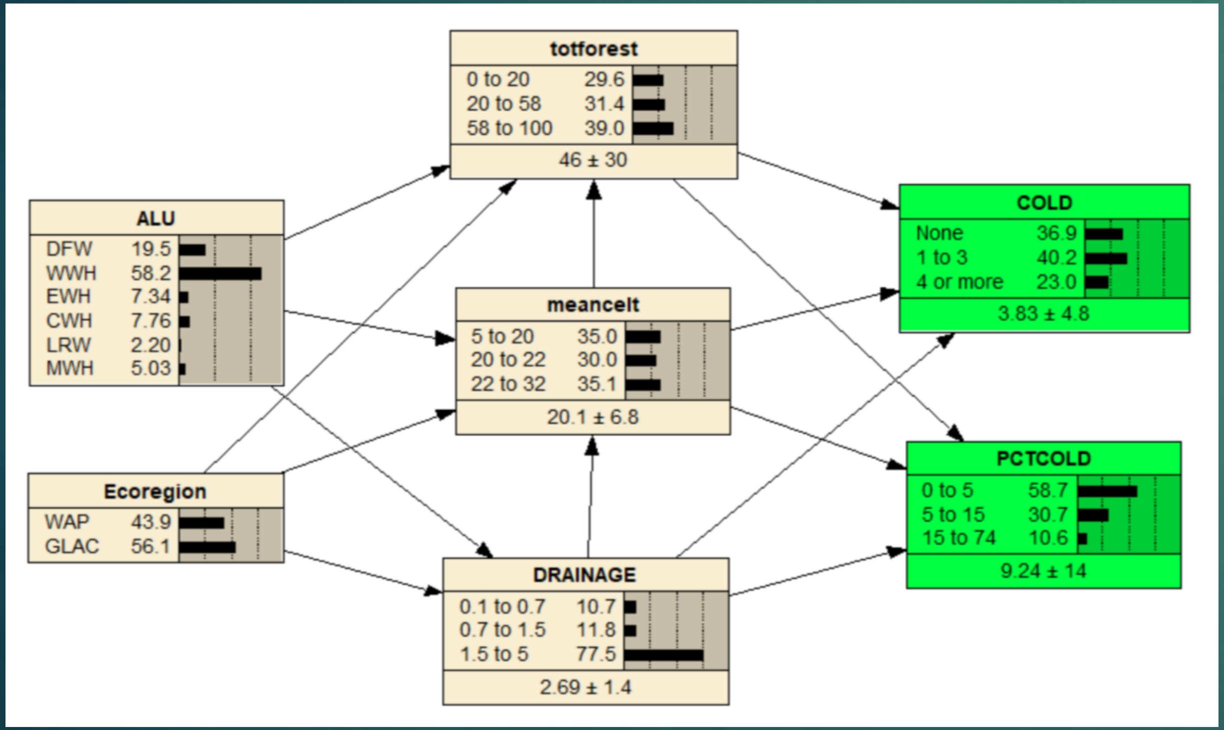


WAP = Western Alleghany Plateau; GLAC = Glaciated

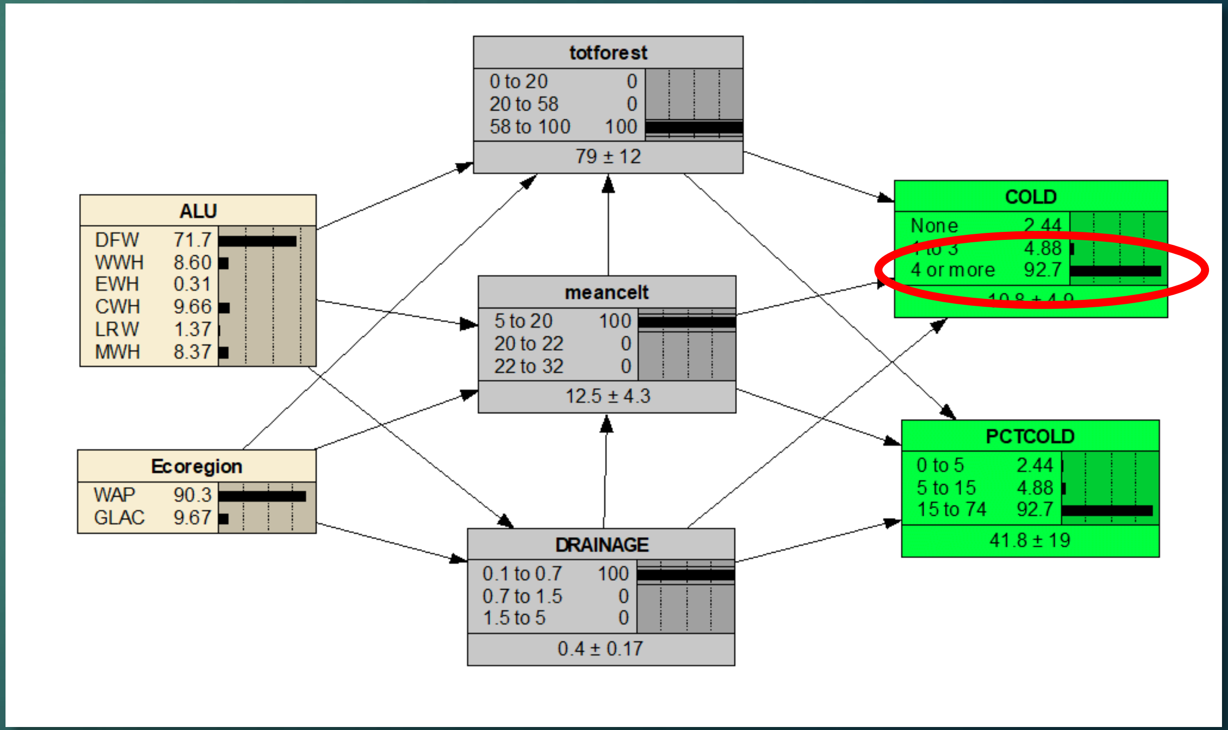
More information in Miltner & McLaughlin (2018)

BN Coldwater Macroinvertebrate Model (Ohio)

Calibrated Model



Model Scenario



WAP = Western Alleghany Plateau; GLAC = Glaciated

More information in Miltner & McLaughlin (2018)

Sensitivity Analysis

- Mutual Information used to compare the relative influence of predictor variables for binned variables
  - Indicates how much having knowledge of one variable reduces the uncertainty in the value of another (equals 0 for independent variables)

Table 4
Sensitivity analysis results by ecoregion for forest area, mean temperature, and drainage area.

Predictor	Ecoregions		
	Combined	WAP	GLAC
	Mutual information (%)	Mutual information (%)	Mutual information (%)
Forest area	9.5	2.4	8.8
Drainage area	4.4	6.8	2.3
Mean temperature	4	8.8	3.1

WAP = Western Alleghany Plateau; GLAC = Glaciated
From Miltner & McLaughlin (2018)

CWH Model Performance:
Model-Predicted Bin Vs. Test Data

K-fold validation using 80/20 train/test

Avg **Correct** Prediction >> 103/191 >> **54%**

Predicted			Total	Actual
Zero	**1 - 3**	**4 or more**		
44 (39-49)	23 (18-28)	4 (0-8)	71	**Zero**
24 (20-29)	40 (34-47)	12 (2-21)	76	**1 – 3**
3 (1-5)	22 (8-35)	19 (9-29)	44	**4 or more**
71	85	35	191	**Total**

Avg **"Worst"** Prediction >> 7/191 = **4%**

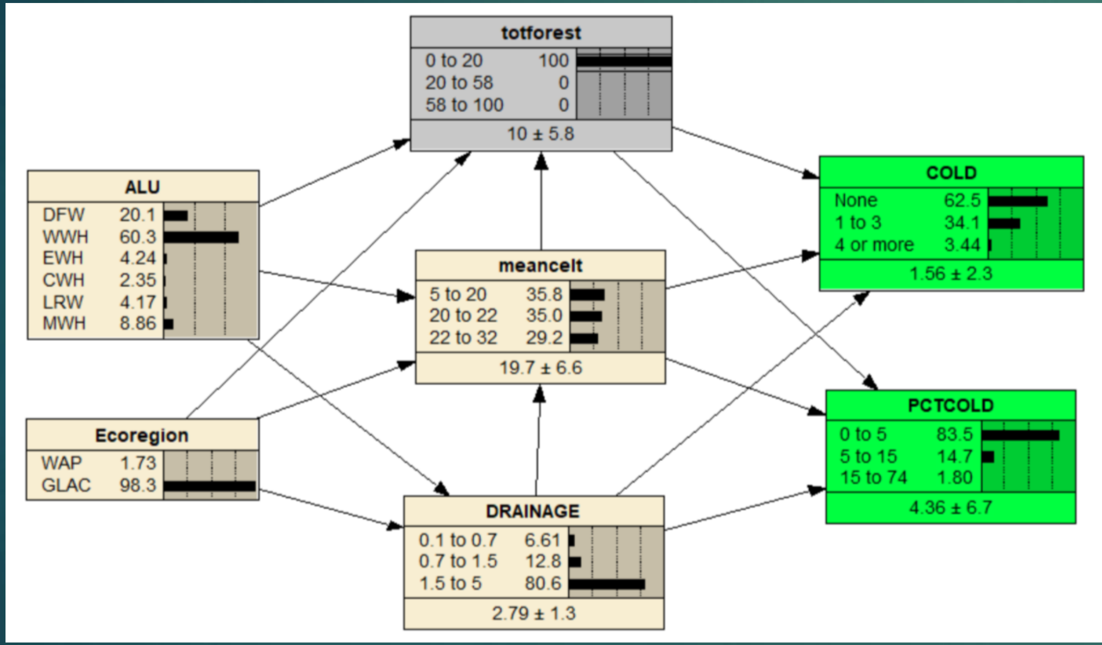
Predicted			Total	Actual
Zero	**1 - 3**	**4 or more**		
44 (39-49)	23 (18-28)	4 (0-8)	71	**Zero**
24 (20-29)	40 (34-47)	12 (2-21)	76	**1 – 3**
3 (1-5)	22 (8-35)	19 (9-29)	44	**4 or more**
71	85	35	191	**Total**

Spherical payoff = 0.66 (1.0 is best)

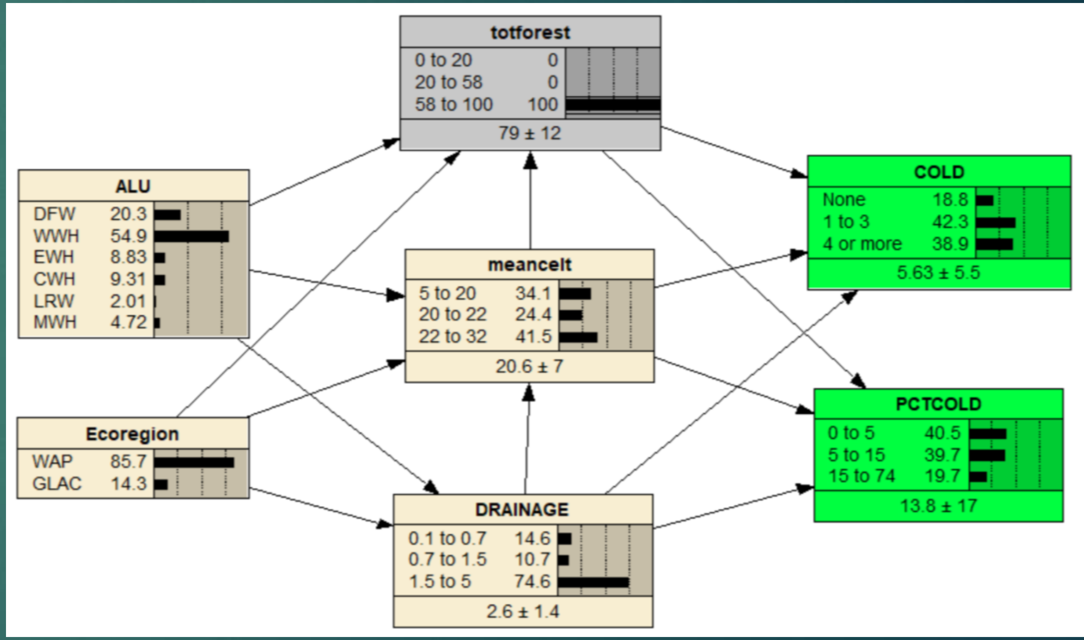
See Miltner & McLaughlin (2018)

Two Scenarios – The Influence of Forest Cover

- Less forest…fewer cold taxa



- More forest…more cold taxa



One take-away: The model says that on average, the chances of finding 4 or more cold taxa are 3.4% when forest cover is 0-20% and 38.9% when forest cover exceeds 58%.

BN Models as a Decision Support Tool: Questions Revisited

Common Management Questions	Relevant BN Model Characteristics
What's our goal, and are we meeting it?	Numeric environmental response targets/thresholds represented as bin boundaries
If not, what action should we take?	Conceptual models, causal inference, management options, sensitivity analyses
How confident are we that the system will respond as we think?	Scenario probabilities, model performance measures, updateable in adaptive management approach
Can stakeholders clearly see what actions are being recommended and why?	Graphic display of model components, interactive model scenarios with a single mouse click, transparent response variable targets, probability-based output

Acknowledgements

Robert Miltner, Kenneth Reckhow, National Council for Air and Stream Improvement, Inc.

Questions?

References:

Marcot, B., Steventon, J., Sutherland, G., McCann, R., 2006. Guidelines for developing and updating Bayesian belief networks applied to ecological modeling and conservation. Can. J. For. Res. 36, 3063–3074.

McLaughlin, D.B., K.H. Reckhow. 2017. A Bayesian Network Assessment of Macroinvertebrate Responses to Nutrients and Other Factors in Streams of the Eastern Corn Belt Plains, Ohio, USA. Ecological Modelling. 345: 21-29. DOI: 10.1016/j.ecolmodel.2016.12.004.

Miltner, R., D.B. McLaughlin. 2018. Management of headwaters based on macroinvertebrate assemblages and environmental attributes. Science of the Total Environment 650: 438–451.

KIESER & ASSOCIATES
ENVIRONMENTAL SCIENCE & ENGINEERING